▪——はじめに

本書は、2010年に初版発行された『プログラマのための文字コード技術入門』の改訂版です。初版は幸いにも多くのご支持をいただき、6回の増刷を重ね、発行された年の東京の大型書店で発表された年間ランキングのコンピュータ書部門で第4位に入りました。

この改訂版では、初版の骨格はそのままに、この9年間の変化を反映して古くなった記述を改め、規格や文字政策、プログラミング言語・API等の最新版に追随するよう全面的に加筆修正しています。

この前書きに目を通している方には、初版を読み終えていて、改訂版を読むべきかを思案中の方もいるでしょう。文字コードの適用、仕様、原理原則の3つに分けて考えると、この9年間で、適用をめぐってはプログラミング言語など変更がありましたが、文字コードの仕様の基本的な部分には大きな修正はありません。原理原則はまったく変わりません。したがって、文字コードというものの概要を知ることが目的の場合は、初版の知識だけでも大体のところは問題ないでしょう。一方、新たに追加された仕様など、知識を最新版へとアップデートしたい方にはこの改訂版が役立つでしょう。

文字コードはしばしば論争の種となるテーマですが、本書は特定のコード系の使用を推奨したり否定したりするものではありません。代わりに、各コード系の特徴や適した用途等を、言語表記の記号としての文字を符号化する観点から根拠に基づいて説明しています。

本書初版に寄せられた感想で予想外に嬉しかったことは、「おもしろい」という感想が多かったことです。本書は技術書ですから第一義的には技術的に役に立つことが求められますし、実際、役に立ったという感想も少なからずいただきました。一方、おもしろいということには、物事への興味を引き出し、より深い理解へ導く力があると筆者は考えます。本書がおもしろいという評価をいただいているのが、文字の符号化という営みの本質的な側面を本書が彫り出し得たためであるならば、著者として望外の喜びです。この改訂版がより多くの方にお楽しみいただけることを期待します。

2018年12月　矢野 啓介

■——はじめに　※初版前書きより。

文字コードは、ソフトウェア技術者にとって必須の知識です。プログラミングをするうえでも、データフォーマットなどを設計するうえでも必要です。しかし、文字コードの知識を体系的に学ぶ機会というのは滅多にありません。断片的な手がかりを元にWebで検索するなどして調べるといった対応をすることが多いと思います。

本書は、ソフトウェア技術者をおもな対象として、文字コードの基礎知識をなるべく筋道立てて説明しようと試みたものです。前半の第1章から第4章までは、文字コードの基本的な概念から始めて、現在日本で使われているものを中心として各種の文字コードを紹介します。後半の第5章から第8章までは、いわば応用編として、コード変換や判別、インターネットでの扱い、プログラミング言語での扱い、そして典型的なトラブルについて説明します。

さまざまな文字コードを本書では取り上げますが、一概にどれを使うべきといった判断は本書は下しません。どの文字コードもそれなりの理由があって作られたものですから、自分のプロジェクトの置かれた状況に応じて有用なものを選んで使えば良いというのが本書のスタンスです。本書ではそれぞれの文字コードがどのような特徴を持っているか、どのような用途に向くかを説明します。また、本書では説明の根拠となる規格等への参照をできるだけ本文中に記すよう努めました。より詳しく知りたい方は、そうした規格等に直接あたって調べることができます。

本書の内容について、内田明氏、高橋征義氏、武者晶紀氏より有益なご助言をいただきました。深く感謝いたします。もちろん、本書の内容に誤りがあれば著者の責任であることはいうまでもありません。

本書の記述のうち意見にかかわるものは、筆者個人の見解であり、いかなる組織を代表するものでもないことをお断りしておきます。

本書が読者の文字コード理解の一助になれば幸いです。

2010年1月　矢野 啓介

■——第2版改訂における、おもな変更点

第2版改訂に伴い、おもに以下の内容の追加・更新を行いました。

- 2010年の改正常用漢字表に対応
- JIS X 0208とJIS X 0213の2012年改正に対応
- ISO/IEC 10646のUCS-4の変更に対応
- Unicodeに追加された変体仮名の説明
- Unicode 11に基づいた説明
- 本書初版では標準化作業中だったUnicode絵文字を、最新版に基づいた説明に変更
- Webブラウザの動作例を2018年時点で使用されているバージョンに更新
- 紹介するコマンドラインツールを最近よく使われるものに見直し
- HTML5の仕様を反映
- YAMLとJSONについて記述
- Javaの対象バージョンを初版の6から、2018年時点で使用の多い8に変更（一部、9にも対応。ただし、それ以降の版でも通用するでしょう）
- Ruby 1.9以降の説明を2.5に基づいて更新
- Javaのデフォルト文字コードの指定方法の説明を追加
- 文字化けのよくあるパターンを紹介
- Unicodeの標準化異体シーケンス（*Standardized Variation Sequence*）の説明を追加
- WebサイトやメールにおいてUTF-8の使用が増えたことを反映した修正

■——謝辞

今回の改訂にあたり、初版に引き続き、内田明氏、高橋征義氏、武者晶紀氏より再度有益なご助言をいただきました。心より感謝いたします。

第2章 文字コードの変遷 021

第7章 プログラミング言語と文字コード 227

7.1 Java 内部処理をUnicodeで行う 229